The Effects of Snap Decision Making on Episodic Recognition Memory

By

Ellen-Jane Hickey

A report submitted as a partial requirement for the degree of Bachelor of Psychological Science with Honours at the University of Tasmania, 2017

Statement of Sources

I declare that this report is my own original work and that contributions of others have been duly acknowledged.

Signed

………………………......

Date

……………………….

Acknowledgements

First and foremost I would like to thank Professor Andrew Heathcote for being my supervisor for this research project. I am very thankful for his guidance, knowledge, reading and editing of my work, without which I would not have been able to produce this thesis. Secondly, I would like to thank David Elliott for writing the python program that automated my experiment task, and Mathew Gretton for helping me set up the program on the lab computers and fixing any computer problems I encountered. Without their help, it would not have been possible to run my experiment. Thirdly, I would like to thank Gabrielle Hartill for agreeing to participate in a pilot run of my experiment. Finally, I would like to thank my friends and family for supporting me through the whole journey from start to end and especially encouraging me to persist through the stressful parts.

Table of Contents

List of Tables and Figures

The Effects of Snap Decision Making on Episodic Recognition Memory

By

Ellen-Jane Hickey

Word Count: 9958

Abstract

The present study is interested in whether recognition memory errors made from snap decisions can be detected and corrected. Fifty two participants from the University of Tasmania made a two-stage recognition decision to words in a test list. The first decision, was a fast (<1sec) recognition memory (old/new) decision; the second a slower confidence response (high, low, change of mind). Words in the test list were manipulated for frequency (high, low), in the English language, and concreteness (high, low). Analysis of the change of mind responses showed an equal reduction from first response to second in the false alarm rate for both high frequency and low frequency words, and a significantly larger improvement in the hit rate for low frequency words from first response to second. Changed responses had a longer processing time (680ms) than unchanged responses (390ms). A significant word frequency mirror effect was also found, but not a significant word concreteness effect. The results indicate that participants can detect and correct errors made from quick decisions, and have important implications for the ability of decision making models' to predict correct responses, and for reducing the noise in speed-accuracy trade-off experiments (Rabbitt, 1969).

Recognition memory tasks for words have long been used in research to understand the decision-making process. The aim of the research has been to determine the mechanisms underlying recognition based decisions in order to develop theories and to develop models that can predict the probability of making one decision over another. However, a limitation of these models is that they do not account for the fact that a person may make an incorrect decision, realise this, and therefore wish to correct it. In experiments manipulating reaction time, such as speed-accuracy trade-off designs, participants make errors, which often leads to slower reaction times on the immediate subsequent trials (Rabbitt, 1969). This slowing down creates noise in the data. Rabbitt theorised that participants slowed down in order to improve their accuracy. This suggests that participants knew they had made an error, and thus were being more cautious to prevent more errors from occurring. If participants can detect they have made an error, then it should follow that they can correct it. Therefore, if experimental designs for speeded reaction time tasks include a mechanism that allows participants the chance to correct an error, it may reduce the slowing down of responses on subsequent trials, and thus reduce the noise in the data.

In order to determine whether fast errors can be detected and corrected, the present study is interested in the effects of snap decision making on episodic recognition memory for words. Words in the study lists will have either a high or low natural language frequency, and be either high or low in concreteness (i.e., non-abstract and abstract). Participants will be required to make a two-stage decision. The first decision will require them to make snap (i.e., within one second) recognition (old/new) decisions to test words, where an old response is given if the participant thinks the test word was in the study list, and a new response if not. Following this initial decision, participants will make a confidence rating on whether they were highly confident in their initial decision, had low confidence, or would like to change their mind.

Episodic memory is the memory system responsible for storing past information about personally experienced events and related temporal information such as when events happened and where (Tulving, 1972; Tulving, 1984; Tulving, 1985; Tulving, 1995). Recognition is defined as the ability to determine whether a test item has been previously studied in the present context (Norman & Wickelgren, 1969). Tulving (1984) distinguishes recognition from recall, as recognition involves both a sense of familiarity when viewing the stimulus item and an internal cue such as 'was this in the study list?', whereas recall involves only retrieving an item from memory based on the internal cue. In recognition memory tasks, the episodic memory system is engaged as participants must recollect temporal information about the test item in order to recognise it.

Mirror Effects

In recognition-memory tasks for words, high frequency (HF) words produce greater false alarm rates (incorrectly identifying a new word as old) than low frequency (LF) words and; LF words produce greater hit rates (correctly identifying old as old) than HF words. The mirroring of the hit rate for LF words to the false alarm rate for HF words is known as the word frequency mirror effect (WFE; Glanzer & Adams, 1985) and is a robust finding in the literature (e.g., Bridger, Bader & Mecklinger, 2014; Glanzer & Adams, 1990; Malmberg, Steyvers, Stephens & Shiffrin, 2002; Stretch & Wixted, 1998).

A mirror effect also occurs for word concreteness, where high concrete words have a greater hit rate than low concrete words, and the low concrete words are more prone to false alarms (e.g., Glanzer & Adams, 1990; Gorman, 1961; Groninger, 1976; Winograd, Cohen & Baressi, 1976). The hit rate portion of the concreteness mirror effect is believed to be due to the fact that high concrete words can be better encoded during study because images can be attached to them, whereas low concrete words are harder to associate with an image (Paivio, 1969). The false alarm rate portion most likely occurs due to the new low concrete words being similar either orthographically or phonetically to study list words, or because their semantic meanings are similar to study list words, thus they are mistaken as old words (Glanzer & Bowles, 1976; Hirshman & Arndt, 1997; Wixted, 1992).

Gorman (1961) studied word frequency and word concreteness together and found that the mirror effect for word frequency was independent of word concreteness. This means that regardless of how concrete or abstract a word was, LF words still had a greater hit rate than HF words, and HF words still had a greater false alarm rate than LF words. Glanzer and Adams (1990) found that a combined mirror effect only occurred when they enhanced the memory strength for old concrete words during the study phase. They suggested that this strengthening was required in order to increase the difference in effectiveness between word frequency and word concreteness.

Dual Process Models

Studies (e.g., Gronlund & Ratcliff, 1989) involving the recognition of items and item pairs highlight how two processes are required to explain recognition memory decisions, as they show that the recognition of a single item can be made by a quick familiarity based decision, but the recognition of an item pair requires a more elaborative and slower search of memory. Dual process theories propose that recognition memory decisions are made based on two separate processes: familiarity and recollection.

Atkinson and colleagues (e.g., Juola, Fischler, Wood & Atkinson, 1971; Atkinson & Juola 1973) propose a familiarity model of recognition memory that conditionally searches memory for elaborative information. Their model assumes that test items have familiarity values that can be measured on a continuous scale. Target items are assumed to have a higher mean familiarity value than lure (new) items, but the distributions for these may overlap. The subjective familiarity of a test item leads to the decision that the item is old or is new, and this decision is made based on set criterion levels. If a high criterion level is reached then an old response is given, if a low criterion level is not reached a new response is given. The model assumes that these old and new familiarity based decisions are made quickly. However, Atkinson and colleagues propose that if the familiarity value falls somewhere between the low and high criterion levels then a more thorough, and therefore slower, search of memory occurs in order to make a decision. The model also proposes that repetitions of both target and lure items in the test list increases the mean familiarity distributions for each. This therefore increases the probability of a lure item requiring a more thorough search of memory in order to correctly reject it. However, increasing the mean familiarity of target items makes them more recognisable, thus they do not require a more thorough search of memory.

Mandler, Pearlstone and Koopmans (1969) proposed a similar dual process model to Atkinson and colleagues, as their theory assumes that recognition decisions are made based on familiarity alone if a criterion level is reached. However, if that criterion is not reached then a more thorough searching of memory occurs in order to find relative associative or elaborative information to correctly recognise the item, or to reject the item if no supporting evidence is found. In terms of word frequency, Mandler (1980) proposed that the familiarity of a word can be increased incrementally, such that increased exposure to a word makes it more familiar. As HF words are already highly familiar, the increase in familiarity a HF word receives at study is not proportional to the increase in familiarity a LF (and thus unfamiliar) word receives. This means at test old LF words can be easily discriminated from LF lures, as the difference between the familiarity values is large. Old HF words however are harder to discriminate from HF lures, as the difference between the familiarity values is small. Therefore, Mandler propose that the WFE occurs due to differences in the familiarity values between HF and LF words at test. Mandler (1980) further extended his original model to explain that familiarity and recollection for an item initiate together and operate in parallel, but familiarity responses occur faster than recollection responses, thus differentiating his model from Atkinson and colleagues' model.

Jacoby and Dallas (1981) through a series of experiments proposed that recognition memory is based on two concepts: relative perceptual fluency, meaning that familiar items are processed more fluently than unfamiliar ones and; elaboration, meaning that an item can be recognised by the associated information attached to it (e.g., the last context it was encountered). Fluent decisions, they argue, are made faster than elaborative decisions, and both systems can initiate together but work independently in parallel, which resembles other dual process theories that propose familiarity and recollection do the same.

Yonelinas' (1994) dual process model assumes recognition is based on both familiarity and recollection. He described familiarity in terms of signal detection theory. Signal detection theory (SDT; Green & Swets, 1966) proposes that people make recognition decisions based on the strength with which a stimulus item matches their memory for the item. Item strength is measured in terms of a pre-set criterion level, which if reached produces an old response. If this criterion level is not reached then the item is determined to be new as there is not a strong memory for the item. Old items are thought to produce greater familiarity distributions than new items, with the difference between the two distributions being the degree of discriminability (*d'*). Yonelinas described recollection as the remembering of qualitative information about the stimulus item (i.e., contextual information) that can be used to ultimately determine whether the item was previously encountered or not. Yonelinas therefore described recollection as being an all-or-none process that does not have a response criterion. This means that the probability of a recollection response occurring remains fixed. Therefore, Yonelinas also assumed that recollection is independent of the false alarm rate. This theory further predicts that recollection responses should only be given in high confidence. Like previous models, Yonelinas' model assumes that familiarity and recollection initiate in parallel but operate independently, with familiarity responses occurring faster than recollection responses.

In summary, dual process theories are in general agreement that recognition memory decisions are based on familiarity and recollection. They further agree that these two processes initiate together but operate independently, with familiarity responses being made faster than recollection responses.

Likelihood Ratio Models

An alternative explanation to the one provided by dual process theories for recognition memory decisions are likelihood or log-likelihood ratio models. Osth, Dennis and Heathcote (2017) explain that log-likelihood models are a transformation of fixed strength signal detection models (i.e., models that have a fixed lure distribution) into a log-likelihood ratio. The ratio is created by taking a single item's strength and calculating the logarithm of the likelihood that the item is a target and dividing this by the likelihood that the item is a lure. Repeating this for the entire strength distribution of all conditions results in a mirror ordered log-likelihood ratio of distributions. For example, the mirror order for an experiment with two conditions (e.g., weak and strong) would be as follows: the lure distribution mean in the strong condition would be lower than the lure distribution mean in the weak condition, thus resulting in the strong condition having a lower false alarm rate than the weak condition and; the target distribution mean in the strong condition would be above the target distribution mean in the weak condition, thus resulting in the strong condition having a higher hit rate than the weak condition.

Attention/Likelihood theory (Glanzer & Adams, 1990; Glanzer, Adams, Iverson & Kim, 1993) proposes that attention plays a key role during learning (i.e., information encoding) and that likelihood ratios play a key role in making a recognition decision. Glanzer and colleagues theory evolved from strength theories, which postulate that recognition decision are made based on item strength, where a test item has features sampled until either a pre-set criterion is met, resulting in an old response, or is not met, resulting in a new response (e.g., SDT). Attention/Likelihood theory goes one step further to propose that the recognition decision is made based on the evaluation of the likelihood ratio that the item is old or is new given the item’s strength.

Dennis and Humphrey’s (2001) Bind Cue Decide Model of Episodic Memory proposes that recognition of words requires the use of a cue to retrieve the context in which the test word was previously encountered. The theory has three key mechanisms: a binding mechanism, which explains how elements of an episode bind together in episodic memory; the cue, which initiates the retrieval of memories and; the decision rule, which determines the end result (e.g., the old or new decision) and is dependent on the likelihood ratio. Other theories such as Shiffrin and Steyvers’ (1997) Retrieving Effectively from Memory model, and McClelland and Chappell’s (1998) Subjective Likelihood in Memory model also propose that there is a likelihood ratio mechanism behind recognition memory decisions.

When the Word Frequency Mirror Effect Fails

Hoshino (1991) conducted four experiments on recognition memory tasks for words and found that in the first three experiments the WFE was not found, as participants appeared to have a bias to respond old to HF words, thus reducing the hit rate advantage for LF words. However, in the fourth experiment, which used the same word lists as the first three, he had participants first complete a lexical decision task (i.e., identification of word and non-words) to better encode the study list words, and as a result the usual mirror effect was found. Hoshino proposed that people use a common criterion to discriminate HF words from LF words, and that the mirror effect occurs when the familiarity distribution for LF words is much greater than the distribution for HF words. From these findings, Hoshino theorised that the increased false alarm rate for HF words is due to greater confusion in tracing the previous context in which the HF word was encountered, whilst the increased hit rate for LF words is due to a more thorough encoding (i.e., of contextual and temporal information) during study, which reduces any confusion.

Joordens and Hockley (2000) propose that the mirror effect occurs due to LF words being less pre-experimentally familiar, thus leading to a lower false alarm rate compared to HF words which are highly familiar. The lack of familiarity for LF words makes them more memorable, and thus easier to recollect during a test, which results in a hit rate advantage. Due to the hit rate advantage for LF words being tied to recollection, Joordens and Hockley theorised that the hit rate portion of the WFE is variable, as recollection can be manipulated, but the false alarm rate portion of the WFE remains constant. In a series of studies they showed that when the ability for recollection to occur was reduced, the hit rate advantage for LF words was reduced or reversed, whilst the greater false alarm rate for HF words remained the same. One of their studies manipulated response time, such that some trials had unlimited time in which participants could make their old/new recognition decision whilst in others they were required to make this decision within 800ms. They found that in the unlimited time trials, the usual WFE occurred, however in the restricted time trails the hit rate for LF words was reduced, whilst the false alarm rate remained higher for HF words. They found that the hit rate was reduced in this instance due to LF words and HF words receiving about an equal number of old responses.

A reduced or reversed hit rate advantage for LF words with a false alarm rate increase for HF words in the WFE are also found in people with dementia (e.g., Balota, Burgess, Cortese and Adams, 2002). Balota and colleagues theorised that this reduced hit rate advantage for LF words is due to a breakdown in the recollection processes in the brain compared to the processes utilised for familiarity. Specifically, they mentioned that participants with dementia were unable to focus their attention on bringing forth recollective information and therefore responded on a less taxing and more automatic familiarity process. In a second experiment, they were able to demonstrate similar effects in a healthy undergraduate student population by requiring participants to make their responses either very quickly (500ms) or slightly slower (1000ms). When required to make fast responses participants could only rely on familiarity, which was reflected in the results by the reduced hit rate advantage for LF words.

This brief review on the literature on how the WFE is affected by familiarity and recollection provides the building blocks for what to expect in regards to the WFE for snap decisions. The present study requires participants to make fast recognition decisions, which means they will be relying on a sense of familiarity. The literature (e.g., Balota et al, 2002; Hoshino, 1991; Joordens & Hockley, 2000) has shown that fast responding results in a reduced hit rate for LF words, but a relatively intact false alarm rate for HF words. False alarms are not the only type of error that can occur in recognition memory. If an old word is incorrectly identified as new, this is known as a miss. As explained by Snodgrass and Corwin (1988), misses occur due to a failure of the test item memory strength to reach the ‘old’ response threshold. Misses are accounted for in the mirror effect as part of the hit rate, as the sum of the probability of a hit and the probability of a miss is 1.0, therefore if the probability of a hit is known, the probability of a miss can be deduced. If fast responding causes the hit rate to reduce for LF words, then this means the miss rate increases for LF words. In the present study we are interested in whether following a snap decision, participants will engage in a longer processing of their recognition decision and thus engage in recollection to correct their false alarms and misses (i.e., by choosing to change their old/new recognition decision). If this occurs then the usual WFE should be shown, with an increased hit rate (due to the correction of misses) following the second decision.

Decision Making Models

Decision making models address how a decision can be reached, and are important for understanding the process that leads to a change of mind decision. Most decision models that account for both confidence and reaction time rely on the concept of sequential sampling. As described by Smith and Vickers (1988), sequential sampling involves the sampling and accumulation of evidence towards a criterion threshold which occurs across three stages. The first stage involves coding sensory information into meaningful information the brain can use to inform a decision. The second and third stages focus on how evidence is accumulated and the stopping criteria for evidence accumulation. These models are able to describe the relationship between sampling time and performance accuracy, and thus are useful in modelling speed-accuracy trade-off effects (Smith & Vickers, 1988). Some decision models propose that the difference between the stimulus sample and the criterion can be used to inform confidence judgements immediately if the difference is large, whereas other models propose taking multiple samples until the difference is large enough to make a response (Van Zandt, 2000). This process of accumulating evidence to make a simple two-choice decision is the foundation for both random-walk and race models.

Van Zandt (2000) explains that both random-walk and race models assume that evidence arrives at the latest decision processing stages (i.e., stages two and three) in support of one answer over the other, however the processes of storing information is different. The random-walk models have one accumulating counter that is pushed towards one threshold or the other by the accumulating sum of evidence. This means that accumulation of evidence towards one response threshold results in the subtraction of evidence towards the other response threshold. Alternatively, race models have an independent counter for each response, such that evidence for both counters can accumulate simultaneously, but the counter that reaches the response threshold first is the response that is given. Van Zandt further explains that both of these models assume that the accumulation process is noisy, meaning that evidence can be incorrectly identified, and that information integration times can vary, which can impact on the accuracy of the decision. However, this noise can be used to help predict evidence accumulation, such that if a person lowers their response threshold they are likely to respond faster but have more errors, whereas a high response threshold will have a slower but more accurate response. This is because when the response threshold is set high, there is a smaller probability that noise will cause errors in the evidence accumulation process.

Link (1975) and Link and Heath (1975) describe a random walk model in terms of their relative judgement theory. This theory proposes that the accumulation of evidence occurs in a random walk pattern, but the starting point for the evidence accumulation can be anywhere between the two response thresholds. Their model also assumes that the average drift rate (i.e., the distance travelled per unit of time; Ratcliff, 1978) towards either response threshold remains the same.

An example of a race model is Vickers' (1970, 1979) accumulator model. This model assumes that two response counters (e.g., counter A and counter B) continuously and independently gather evidence for or against their respective response (e.g., response A and response B) until one counter reaches its criterion level (threshold). The criterion level for each response counter is variable such that is can be subjectively changed depending on a person's motivation, effort, bias or some other factor. This means, in contrast to relative judgement theory, instead of the start point being adjusted under bias, the threshold for a response in this model is lowered under bias. Vickers (1979) developed the balance of evidence (BoE) hypothesis to accompany his accumulator model in order to explain differences in confidence. The BoE hypothesis explains that a decision will be made with high confidence when there is a large difference between the amounts of evidence accumulated by each response counter. If this difference is small, the decision given will be made with low confidence.

Change of Mind

A lay conceptualisation of how a decision can be changed is the parallel process committee decision model (Rabbitt, Cumming & Vyas, 1978; Rabbitt & Rodgers, 1977). This model proposes that independent committee members cast their ‘vote’ towards a response, and that a decision can be based on a few members’ votes. When all members have voted, it can be determined if the decision was correct (i.e., the slower votes agreed with the faster ones), or if there was an error made (i.e., slower votes disagreed). According to this model errors are more likely to lead to a change of mind. Rabbitt and Vyas (1981) tested this theory in two sensory discrimination tasks that required quick responses. They found that as discriminability became harder participants made more errors and had lower confidence in their decisions. However, they found that when participants had longer exposure to the stimuli in these difficult trials, they were better at detecting that they had made an error and thus correcting it, than when the difficult trials were paired with short exposure times.

Change of mind has also been studied in sensorimotor tasks. A study by Resulaj, Kiani, Wolpert and Shadlen (2009) required participants to view a random dot motion stimulus and move a handle with their hand to one of the two targets they believed the dots were moving towards. In some trials they observed that participants changed direction part way to a target. They theorised that this was due to the initial hand movement being based on partial accumulation of evidence, and that the change of direction occurred due to a full processing of all information that resulted in a disagreement with the initial decision. This finding has since been replicated in more recent studies (e.g., Burk, Ingram, Franklin, Shadlen & Wolpert, 2014; van den Berg et al. 2016).

Van Zandt and Maldonado-Molina (2004) examined change of mind decisions under bias in a fast two-stage recognition memory task for words. Using Vickers' (1979) BoE hypothesis, they predicted that a bias to respond old would cause participants to change correct responses to high bias items and incorrect responses to low bias items more frequently than any other items, as at the time of their confidence decision the alternative response should have accumulated more evidence than the response given. However, they found that a bias to response old equally affected all item conditions, and that the frequency in which changes of mind occurred was due to primary decision accuracy and not bias (i.e., participants only changed their responses when they had made an error). They therefore theorised that after the primary decision, evidence for each response continues to accumulate towards a second response threshold. When one counter reaches its second threshold, confidence is determined by the difference between the amounts of evidence accumulated for each counter. If more evidence has accumulated for the alternative counter then the original response is determined to be incorrect and therefore changed.

Van Zandt and Maldonado-Molina (2004) also investigated whether there would be significant differences between the reaction times for consistent responses (i.e., the confidence rating reaffirms the old/new recognition decision) and inconsistent responses (i.e., the confidence rating changes the initial recognition decision). They found that consistent responses were made significantly faster than inconsistent responses, which provides evidence that changed responses incur longer processing times, and is a result that the present study is interested in replicating.

Curran, DeBuse and Leynes (2007), using Van Zandt and Maldonado-Molina's theory and design, also studied change of mind under bias in a recognition memory task for words, and found that participants changed their mind when their first response resulted in a miss, or to correct for another type of error (e.g., false alarms).

Aim and Hypotheses

The present study is concerned with understanding the effects of making a snap decision on episodic recognition memory, with particular interest in whether making snap decisions causes errors which can be detected and corrected via a change of mind response. If participants are found to be able to detect and correct errors made under time pressure, this could have important implications for reducing noise in the data for speed-accuracy trade off experiments, and for improving the probability with which decision-making models can predict a correct response. The present study therefore aims to investigate the effect of time pressure on recognition memory of words of varying frequency (high and low) and concreteness (high and low) by requiring participants to make a fast two-stage recognition decision. Several hypothesis have been devised from previous research findings in order to test this aim:

1. Mirror effects will be observed independently for both word frequency and for word concreteness, with the hit rate portion of the mirror effects being larger for the second decision than the first decision.
2. Recollection will correct misleading familiarity information for new items, meaning that there will be a greater rate of change of mind for high frequency new words (where familiarity is misleading) than for low frequency new words (where it is not).
3. Recollection will correct familiarity based misses, as there will be a greater rate of change of mind for low frequency old words (where familiarity is misleading) than for high frequency old words (where it is not).
4. Van Zandt and Maldonado-Molina's (2004) finding that consistent responses are made faster than inconsistent responses will be replicated.

Method

Participants

Fifty-two participants from the University of Tasmania participated in the study. Twelve participant's data files were not used in the analysis of the study. One participant's data file was not saved due to computer error; two participants were removed for having more than 10% of non-responses; five were removed for having a final accuracy of less than 55% and; four were removed for having less than 10 'change' responses. The analysis therefore consisted of 40 participants. First year undergraduate psychology students received one hour of course credit for participating, whilst other university students had the option to receive $15 in remuneration for their time.

Materials

The total experimental item set was made up of 2,069 nouns and verbs, of which 1,536 were used as experimental trials and 533 were used as practice trials. Items were rated for word frequency (min=1, max=314232, median=469) and contextual diversity (min=1, max=8363, median=294), according to the subtitles lexicon of American English Brysbaert and New (2009) norms. Word concreteness ratings followed the Medical Research Council psycholinguistic database (MRC; Coltheart, 1981) and referred to the rating for imageability scores (min=183, max=667, median=506). The length of each word was between four and seven letters.

A total of 866 words from the original word pool were randomly selected to use for each participant's test session. Twenty-four of these words were used in the practice trial, and 64 were used as buffer words throughout the experimental trials. Therefore, the experiment consisted of 778 test words. Seventeen sub-lists were made, allowing for one practice cycle and 16 recognition-memory cycles. Words within each list were randomly assigned and counterbalanced for word frequency, contextual diversity, and concreteness. The practice cycle consisted of a study list of 12 words, and a test list of 24 words (i.e., the 12 study list words and 12 new words). The experimental cycles comprised study lists of 28 words and test lists of 50 words. There were four foil (i.e., buffer) words in each study list, two at the beginning and two at the end, with only two of these words randomly selected to be in the test list but not analyzed. Foils were included in order to control for primacy and recency effects. Each test list in the experimental trials therefore consisted of 24 target words and 24 new words.

The memory task was automated by a program written in Python (v.26) language and run on an IBM compatible computer with Windows OS, and a QWERTY keyboard. The test instructions were displayed on a 24inch colour monitor with 1920x1080 pixel resolution in Arial font, size 48.

Procedure

Participants were seated in front of a computer and read through the experiment instructions at their own pace. Participants undertook key pressing training in order to familiarise themselves with the response keys. The required keys to press alternated between participants, such that half the participants pressed the old/new keys using their right hand, and half pressed the old/new key using their left hand. This means that half the participants used their right hand to press the ';' key for an old recognition decision and the '.' key for a new decision, and used their left hand to press the 'z' for high confidence; 'x' for low confidence and; 'c' for change of mind. The other half of participants used their left hand to press the 'a' for an old recognition decision and the 'z' for a new decision, and used their right hand to press the 'm' for high confidence; ',' for low confidence and; '.' for change of mind. Participants were encouraged to keep their middle and index fingers of one hand on the old and new keys, and their index, middle, and ring fingers of the other hand on the confidence rating keys.

Following the key training exercise, participants completed a practice trial. Study list words appeared for one second, with a one second break between word appearances.

Following the last word in the study list there was a 15 second interval, in which participants were reminded on how to respond to words in the test list. The test list followed this short break, with words again appearing one at a time. Participants had one second to make their old/new recognition decision before being asked for their confidence response. Participants were allowed up to five seconds to make their confidence response. The next word in the test list appeared after participants had either made their confidence response, or they failed to make a response in six seconds. If participants took 750ms or longer to make their old/new decision, the message 'TOO SLOW' appeared on the screen in red font. The message 'TOO FAST' appeared in red font if the participant responded in less than 250ms. If the participant failed to make any response in six seconds, the message 'TIME LIMIT EXCEEDED, NO RESPONSE RECORDED!' appeared in red font. Participants were encouraged to avoid receiving all three of these messages. The experimenter remained in the room until the participant had completed the practice trail in order to answer any questions or resolve any confusion about the task. The experimental trials were identical to the practice trial, with the exception of word list lengths. At the end of each cycle, participants received feedback on their accuracy (out of 100%). At this point, participants were encouraged to improve their accuracy in the next cycle by choosing the change option when they were unsure of their old/new decision or when they knew they had made a mistake.

In order to encourage accuracy and speed in their initial response, as well as using the change of mind option and having well calibrated confidence (i.e., being more accurate for high than low confidence), participants received points following each old/new, high/low/change decision they made. The points system worked as follows: for a correct old/new decision participants received 1000 minus their reaction time in milliseconds in points. They earned an additional 500 points if they responded with high confidence, or 100 points if they responded with low confidence. If they responded with change of mind, they received zero points. If a participant made an incorrect old/new response, they received zero points, if they followed this decision with a high confidence rating they had 500 points subtracted from their score, or 100 points subtracted if they responded with low confidence. If they chose to change their mind they received 500 points. Participants were encouraged to gain as many points as they could.

Design and Data Analysis

The study is a 2(word frequency: high, low) x 2(word concreteness: high, low) within-subjects design. The dependent variables are reaction time (ms), accuracy, and the confidence rating response (high confidence, low confidence, change of mind). Generalised (binomial probit) linear mixed models were used to analyse the hit and false alarm rates, and change of mind rates were analysed using the R package *lme4* (Bates, Maechler, Bolker & Walker, 2015), and inferences were conducted via Wald χ^2 tests with type III sums of squares as implemented by the *car* package (Fox, Friendly, & Weisberg, 2013).

A linear mixed-effect (LME) model is an extension of the standard linear model. These models contain both fixed-effects and random-effects, hence the name ‘mixed-effect’. Participants are almost always treated as a random effect but experiments such as the present one have another source of random variation from the different word stimuli used. Freeman, Heathcote, Chalmers and Hockley (2010) advocate using both subject and item random effects in the analysis of recognition memory. LME models are run in a hierarchical manner, such that the simplest model (e.g., only containing fixed-effects) is compared to a more complex model or models (i.e., containing one or more random-effects in addition to the fixed-effects) in order to see which model better fits the data. Model fit can be determined through either the Akaike Information Criterion (AIC; Myung & Pitt, 1997), which is calculated from the number of parameters in the model, or the Bayesian Information Criterion (BIC; Myung & Pitt, 1997), which is calculated from the number of parameters in the model and the total number of observations used to fit the model (Pinheiro & Bates, 2000). In both cases, a smaller number indicates a better model fit (Pinheiro & Bates, 2000).

Generalised linear models (GLM) are the best type of model to use when data is non-normally distributed, as it utilises link functions to take account of distribution shape (McCullagh & Nelder, 1989). Generalised linear mixed-effects models integrate both GLM and LME models, making them a powerful analysis tool that can analyse non-normally distributed data that also has random effects (McCullagh & Nelder, 1989). For the present analyses, where the data are binary, a binomial error model is natural. We assumed the binary recognition responses could be approximated by an equal-variance signal detection model (Heathcote, 2003[1]) and so used a probit link function (Rouder & Lu, 2005), which is equivalent to an analysis of results on the inverse cumulative normal (i.e., z) scale. The analysis of response choices was carried out in terms of the probability of responding old (i.e., that the test word was studied), which constitutes the hit rate (HR) for studied (old) items and the false alarm rate (FAR) for lure (new) items. In signal detection theory recognition ability is measured by *d'* = z(HR)-z(FAR); in the LME analysis of the probability of responding old, reported effects of a factor on *d'* correspond to an interaction with a factor representing the response type (new vs. old).

Results

In the following analyses, the simple models (containing only a random-subject effect) were compared to the more complex models, which contained both random-subject and random-item effects. Following Freeman et al.'s (2010) advice the more complex model

---

[1] ROC analysis reported in this paper and others supports an unequal variance model, but this is not available within the generalized linear model family and so cannot be used with standard linear mixed models packages. Hence, we used the equal variance approximation.

containing both random-subject and random-item effects was always interpreted. The model fits have not been reported in-text but can be found in Appendix D. For ease of reading, the analysis is broken down into four parts. The first part analyses data on the first response (i.e., the old/new response). The second part analyses the data for the second response (i.e., the old/new response taking into account the confidence response high, low, or change). The third part analyses data from parts one and two combined. The fourth part analyses the effect of response consistency on reaction time.

First Response

Participant's old/new responses to words of high and low frequency and high and low concreteness were analysed using an ANOVA. There was a significant interaction between response type (old/new) and word frequency, $\chi^2(1)=30.2$, $p<.001$, indicating that *d'* for LF words (0.58) was significantly larger than *d'* for HF words (0.43). The mean percentage of old responses broken down by word frequency and word concreteness conditions are displayed in Table 1. If the probability of a hit rate for LF words mirrors the probability of a false alarm rate for HF words, then a word frequency mirror effect is present (Glanzer & Adams, 1990). Table 1 shows that the probability of responding old to old items (hit rate) is higher for LF words than HF words, and the probability of responding old to lure items (false alarm rate) is higher for HF words than LF words, thus the expected word frequency mirror effect is shown. The hit and false alarm rates for word frequency were analysed in an ANOVA, which shows that the hit rate for old LF words is significantly larger than the hit rate for old HF words, $\chi^2(1)=21.9$, $p<.001$ and; the false alarm rate for new HF words is significant larger than the false alarm rate for new LF words, $\chi^2(1)=8.54$, $p=.003$.

A concreteness mirror effect is shown when the probability of a hit rate is higher for high concrete (HC) words, and the false alarm rate is higher for low concrete (LC) words (Glanzer & Adams, 1990). Table 1 shows the hit rate is higher for HC words compared to LC words, but the false alarm rate is equal for LC and HC words, thus only the hit rate portion of the expected word concreteness mirror effect is present. The hit and false alarm rates for word concreteness were analysed in an ANOVA, which shows that the hit rate for old HC words was significantly larger than the hit rate for old LC words, $\chi^2(1)=10.5$, $p=.001$, but there was not a significant difference between the false alarm rates for new LC words and new HC words ($p=0.69$).

Table 1

First Response Mean Probabilities (%) of Responding Old to New and Old Items Under Different Word Frequency and Word Concreteness Conditions

Stimulus Item	HF	LF	HC	LC
New	41	38	41	40
	40	38	38	38
Old	59	62	59	56
	56	60	62	60

Linear mixed models were used to analyse reaction time (RT) data. When responses are made under time pressure, errors (i.e., false alarms and misses) are made faster than correct responses (Ratcliff & Rouder, 1998). The effect of first response correctness (C1) on RT was analysed in an ANOVA. The expected result was shown as there was a significant main effect for C1, $\chi^2(1)=134.1$, $p<.001$, indicating that errors (M=540ms) were made significantly faster than correct responses (M=580ms).

Second Response

Initial analysis of the second response (which included the previously excluded four participants who had less than 10 change responses), suggested that some participants made their second response very quickly, indicating that they did not follow instructions to use the second response as an opportunity to improve their accuracy. In order to quantify this effect we analysed the confidence response RTs with change in accuracy between first and second responses using *t* tests. There was a significant moderate negative correlation between the proportion of fast confidence responses, where fast was defined as less than 200ms, and change in accuracy, *t*(42), *p*<.001, *r*=-.59, indicating that those who made their confidence response quickly were more likely to have a lesser increase in final accuracy. Of those who had many fast confidence response RTs, the confidence response they nearly always gave was ‘high’. There was a significant moderate positive correlation between mean confidence RT and change in accuracy, *t*(42), *p*=.002, *r*=.44, indicating that there was a trend for improved accuracy with longer confidence response RTs. Figure 1 shows these trends graphically.

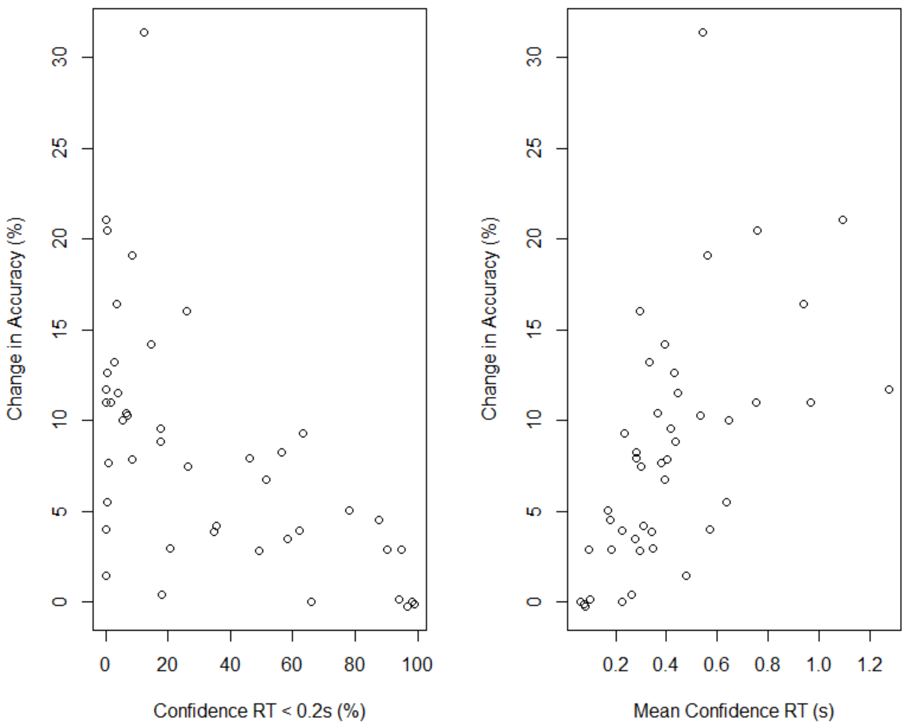

Figure 1. The relationship between fast confidence responders and change in accuracy (left hand side) and; the relationship between the mean confidence response RT and change in accuracy (right hand side).

Participant's old/new responses to words of high and low frequency and high and low concreteness were analysed using an ANOVA. There was a significant interaction between response type and word frequency, $\chi^2(1)=84.8$, $p<.001$, indicating that *d'* for LF words (1.13) was significantly larger than *d'* for HF words (0.86). There was also a significant interaction between response type and word concreteness, $\chi^2(1)=12.3$, $p<.001$, indicating that *d'* for HC words (1.05) was significantly larger than *d'* for LC words (0.94). The mean probability (%) of responding old in the word frequency and word concreteness conditions are displayed in Table 2. The hit rate is higher for old LF words than old HF words, and the false alarm rate is higher for new HF words than new LF words, thus the expected word frequency mirror effect is present. The hit and false alarm rates for word frequency were analysed in an ANOVA, which shows that the hit rate for old LF words is significantly larger than the hit rate for old HF words, $\chi^2(1)=68.9$, $p<.001$ and; the false alarm rate for new HF words is significant larger than the false alarm rate for new LF words, $\chi^2(1)=16.2$, $p<.001$.

Table 2 shows the hit rate is higher for HC words compared to LC words, but the false alarm rate is equal for LC and HC words, thus only the hit rate portion of the expected word concreteness mirror effect is present. The hit and false alarm rates for word concreteness were analysed in an ANOVA, which shows that the hit rate for old HC words was significantly larger than the hit rate for old LC words, $\chi^2(1)=13.8$, $p<.001$, but there was not a significant difference between the false alarm rate for new LC words and HC words ($p=0.32$).

Linear mixed models were used to analyse RT data. The second response deliberately allowed participants longer to make their decision, thus placing an emphasis on considering the accuracy with which they had made their first response. When accuracy is emphasised, correct responses are made faster than incorrect responses (Swensson, 1972). The effect of second response correctness (C2) on RT was analysed in an ANOVA. The expected result was found as there was a significant main effect for C2, $\chi^2(1)=48.6$, $p<.001$, indicating that correct responses (M=444ms) were made significantly faster than mistakes (M=454ms).

Table 2

Second Response Mean Probabilities (%) of Responding Old to New and Old Items Under Different Word Frequency and Word Concreteness Conditions

Stimulus Item	HF	LF	HC	LC
New	31	27	31	31
	31	28	27	28
Old	65	72	65	63
	63	69	72	69

Combined Analysis

Part three of the analysis combined data from the first and second response and so includes an additional variable R12 with two levels: response 1 (i.e., first response data), and response 2 (i.e., second response data). Combining the first and second responses in the one analysis is necessary for determining whether changes of mind improved participants accuracy (i.e., whether participants corrected their false alarms and misses).

R12 and participant's old/new responses words of high and low frequency and high and low concreteness were analysed using an ANOVA. There was a significant three way interaction between R12, response type, and word frequency, $\chi^2(1)=9.34$, $p=.002$, indicating that the mean *d'* difference between LF and HF words for response 2 (0.28) was significantly larger than the mean *d'* difference between LF and HF words for response 1 (0.15).

ANOVAs were used to examine the hit and false alarm rates for R12. The analysis was of results on the z-scale, but here they have been transformed into percentages (%) for reporting trends. There was a significant interaction for hit rate between R12 and word frequency, $\chi^2(1)=10.9$, $p<.001$, indicating that the mean probability of a hit for HF words is larger for response 2 (64%) than response 1 (58%), and the mean probability of a hit for LF words is larger for response 2 (71%) than response 1 (61%). There was also a significant main effect for R12, $\chi^2(1)=231.4$, $p<.001$, indicating that the difference between the hit rate for HF and LF words was higher for response 2 (7%) than response 1 (4%). A significant main effect for word frequency was also found, $\chi^2(1)=54.5$, $p<.001$, which indicates that the difference in hit rates between response 1 and response 2 for LF words (10%) was greater than the difference between hit rates between response 1 and response 2 for HF words (6%).

The mean false alarm rate (%) for HF and LF words in both response 1 and response 2 are as follows: response 1 (HF=41, LF=38), response 2 (HF=31, LF=28). There was a significant main effect for false alarms for R12, $\chi^2(1)=389.8$, $p<.001$, indicating that the difference between the false alarm rate for new HF and new LF words was slightly larger in response 2 (3%) than response 1 (2%). There was also a significant main effect for false alarms for word frequency, $\chi^2(1)=14.64$, $p<.001$, indicating that the difference between the false alarm rate for response 1 and response 2 for new LF words (11%) was larger than the difference for HF words (10%).

Response Consistency

This final part of the analysis concerned the reaction time for consistent responses (i.e., the confidence response reinforces the old/new decision) and inconsistent responses (i.e., the confidence response changed the first response). It is expected that consistent responses will be faster than inconsistent ones (see Van Zandt & Maldonado-Molina, 2004). The analysis was conducted using a linear mixed effects model with the added variable 'consistency'.

The confidence response RTs, second response decision, and consistency were analysed using an ANOVA. There was a significant main effect for consistency, $\chi^2(1)=4040.9$, $p<.001$, indicating that inconsistent responses (M=680ms) were significantly slower than consistent responses (M=390ms). There was a significant interaction between response type and consistency, $\chi^2(1)=31.9$, $p<.001$, indicating that the difference in mean RT between consistent and inconsistent new responses was 224ms, with consistent responses being faster than inconsistent ones and; the difference in mean RT between consistent and inconsistent old responses was 343ms, with consistent response being faster than inconsistent ones.

Discussion

The present study aimed to investigate the effects of time pressure on recognition memory, with particular interest in whether participants could detect and correct their errors. Hypothesis 2, hypothesis 3 and hypothesis 4 were of primary interest, as these specifically looked at the effects of change of mind. Hypothesis 2 predicted that participants would use recollection to correct familiarity based false alarms for new HF words more than new LF words. Hypothesis 3 predicted that participants would use recollection to correct familiarity based misses for old LF words more than old HF words. Hypothesis 4 predicted that consistent responses would be made faster than inconsistent responses, thus replicating Van Zandt and Maldonado-Molina's (2004) finding. The results provided some support for hypothesis 2, and fully supported hypothesis 3 and hypothesis 4. The remaining hypothesis was of secondary interest and acted more as a manipulation check. Hypothesis 1 predicted that there would be independent mirror effects for both word frequency and word concreteness, with an increased hit rate portion of the mirror effect following the second response. The results provided partial support for this hypothesis.

Hypothesis 2 was only partially supported because the expected reduction in the false alarm rate for new HF words was not found. The results showed that the false alarm rate was reduced from the first response to the second response, indicating that participants were using the change of mind option to correct their false alarms. However, the reduction in the false alarm rate for new HF and LF words was about equal, which did not support the hypothesis. LF words are less familiar than HF words, thus they usually have fewer false alarms due to their lower familiarity (Hoshino, 1991; Joordens & Hockley, 2000; Mandler, 1980). According to dual process theories (e.g., Jacoby & Dallas, 1981; Mandler, 1980; Yonelinas, 1994) recollection of more elaborative information at the time of the second response should have weakened the item strength of the lure word, thus allowing participants to realise their mistake. Due to the higher familiarity of HF words, recollection should have had more of an impact on the false alarm rate for HF words than LF words. The pattern of results found in the present experiment do not confirm the dual process predictions, nor do they falsify them. Therefore, more data should be collected in a more powerful design to see if the null difference replicates.

Likelihood ratio theories (e.g., Dennis & Humphrey's, 2001; Glanzer et al., 1993; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997) propose that recognition decisions are based on the likelihood ratio of an item being old or new given the strength of the item. If participants in the current study were basing their decisions on a likelihood ratio instead of on the familiarity distributions of the test items, this may explain why the reduction of the false alarm rate for LF words and HF words was about equal. Likelihood ratio models do not currently account for change of mind decisions. However, if the models were elaborated it would be possible to test whether they are adequate in explaining the results found in the present study.

Hypothesis 3 was operationalised in terms of the hit rate, as the hit rate portion of the word frequency mirror effect encompasses the probability of the miss rate (see Snodgrass & Corwin, 1988). Thus, it was predicted that if the difference in hit rate between old LF and old HF words for response 2 was greater than the difference in hit rate between old LF and old HF words in response 1, then participants were increasing their word hit rate by decreasing the respective miss rate (i.e., by correcting their misses). The hit rate was found to be larger for response 2 than response 1 and in the predicted direction, indicating that participants were detecting when they 'missed' a word and were able to correct this error by using the change option. This finding is in line with Curran et al.'s (2007) study that also found that participants were correcting their misses as operationalised in terms of the improved hit rate following the second response.

The fourth hypothesis was supported as consistent responses were found to be made significantly faster than inconsistent responses. Therefore the present study successfully replicated Van Zandt and Maldonado-Molina's (2004) finding. The significant interaction between response type and consistency showed that for both old and new responses, responding consistently was always faster than responding inconsistently. These findings imply that decisions that are changed undergo a longer processing time than responses that are not changed.

Hypothesis 1 was partially supported as a word frequency mirror effect (WFE) was found, with an increased hit rate for LF words following the second response, but no significant word concreteness effect was found following either the first or second response. Interestingly, a significant WFE was found for the first response. Balota and colleagues (2002) and Joordens and Hockley (2000) in their studies did not find a significant WFE when their participants were required to make fast recognition decisions. Their experiments, however, required participants to respond even quicker (500ms and 800ms respectively) than in the present experiment (1000ms). Joordens and Hockley theorised that the WFE does not occur under time pressured responding due to there not being enough time for participants to recollect seeing a word before, thus the hit rate for LF words reduces, but the false alarm rate for HF words remains intact as this is not dependent on recollection. Balota and colleagues' experiment compared fast (500ms) responding to slower (1000ms) responding, and found that the WFE was not present in the fast condition but was in the slower condition. As both Balota and colleagues experiment and the present experiment found a WFE for responding within 1000ms, this may indicate that 1000ms is enough time for some recollection to occur for LF words, thus keeping the hit rate higher for LF words than HF words and producing the word frequency mirror effect.

A significant WFE was also present for the second decision, and as predicted the hit rate for old LF words increased. The proportion of false alarms for new HF words had also reduced relative to the first decision. *d'* was larger for both LF and HF words in response 2 compared to response 1, with the mean *d'* difference between LF and HF words in response 2 being larger than the difference between LF and HF words in response 1. Taken together, these results imply that the distributions for LF and HF words moved further apart following the second response, thus meaning that there was reduced overlap between the distributions of HF and LF words, and therefore a reduction in the false alarm rate. This provides further support that participants were using the change of mind option to improve their accuracy.

The lack of a significant word concreteness effect was due to the false alarm rate between HC and LC words not being significantly different. The false alarm rate for LC words most likely occurs due to a lack of distinctiveness of the LC lures from memory of the study list words, meaning that LC lures either appear orthographically or phonetically similar to words that were in the study list, or their semantic meanings are similar to words that were in the study list (Glanzer & Bowles, 1976; Hirshman & Arndt, 1997; Wixted, 1992). Glanzer and Adams (1990) explain that the word concreteness mirror effect may not occur when words within the test lists are too similar in their concreteness rating. Although the words used in the present experiment were rated on concreteness (ranging from 183-667 according to the MRC; Coltheart, 1981), it is possible that when they were randomly allocated to the test lists they did not differ enough in their concreteness, thus there was not a clear enough distinction between HC lure words and LC lure words to produce the false alarm rate portion of the concreteness mirror effect.

There were two key limitations in the present study that impacted the results. The first limitation was that participants often made their second response quite fast. This is an issue because it may have impacted participant's ability to use recollection to inform their decisions. Yonelinas and Jacoby (1994) argue that recollection responses on average peak approximately between 800ms and 1100ms. Participants in the present study made their first response (i.e., old/new recognition decision) on average in about 560ms. Combining this with a confidence response reaction time of 200ms or less results in a total processing time of approximately 700-800ms, and therefore places the final decision outside of the timeframe in which recollection peaks according to Yonelinas and Jacoby. If participants had taken more time to consider their confidence response, there may have been larger improvements in final accuracy (i.e., by correcting more false alarms and misses) than was seen. Future replications of this experimental design should therefore consider preventing participants from being able to make their confidence response quickly, thus forcing them to slow down and think about their response. The second limitation is that participants generally tended to not use the low confidence option. This means that analyses, such as receiver operating characteristics (which is a plot of the relationship between the hit and false alarm rates for different confidence response ratings; Yonelinas, 1997), could not be conducted on the confidence responses. Future replications of this design should therefore also consider including a mechanism to encourage participants to use all of the response buttons equally. For example, in the present study participants were encouraged at the end of each cycle to improve their accuracy by using the change option when they were unsure of their response. This message could be altered to reinforce using all of the confidence response options.

The present study has shown that following a snap decision, participants are able to detect when they have made an error such as a false alarm or a miss and correct it, with change of mind responses being made on average within one second of making the initial response. These findings have important implications for speeded response experimental designs and decision making models. Decision making models should consider including parameters in their equations for the probability of an error being corrected, as this will have an impact on determining the probability of a correct response being made. Speeded response studies should consider including a change of mind option in their experimental designs in order to give participants the opportunity to correct their errors and thus reduce noise in the data that is created by a slowing down of responses following a known error (see Rabbitt, 1969).

In conclusion, the present study was able to show that errors made following a snap recognition decision can be detected and corrected, often within one second of making the error. The results showed that participants were correcting their familiarity based misses and false alarms in order to improve their accuracy. However, the reduction of the false alarm rate in hypothesis 2 was eqaul for new HF and LF words which did not supported the hypothesis and therefore could neither confirm nor falsify the pattern of results predicted by dual process theories. It was recommend that the experiment be replicated with more power in order to see if the null result replicates. It was also proposed that if likelihood ratio models are elaborated to account for change of mind decisions, then they may also be able to explain the pattern of results found in the present study. Nonetheless, the fact that participants could detect and correct their errors has important implications for improving the ability for decision making models to predict correct recognition decisions, and for reducing noise in speed-accuracy trade-off experimental designs.

References

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Fourth International Symposium on Attention and Performance* (pp. 583-612). New York: Academic Press

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The Word-Frequency Mirror Effect in Young, Old, and Early-Stage Alzheimer's Disease: Evidence for Two Processes in Episodic Recognition Performance. *Journal of Memory and Language, 46*, 199-226. doi:10.1006/jmla.2001.2803

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statisitcal Software, 67*, 1-48. doi:10.18637/jss.v067.i01

Bridger, E. K., Bader, R., & Mecklinger, A. (2014). More ways than one: ERPs reveal multiple familiarity signals in the word frequency mirror effect. *Neuropsychologia, 57*, 179-190. doi:10.1016/j.neuropsychologia.2014.03.007

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977 990. doi:10.3758/BRM.41.4.977

Burk, D., Ingram, J. N., Franklin, D. W., Shadlen, M. N., & Wolpert, D. M. (2014). Motor effort alters changes of mind in sensorimotor decision making. *PLoS One, 9*(3), e92681. doi:10.1371/journal.pone.0092681

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A, 33*(4), 497-505. doi:10.1080/14640748108400805

Curran, T., DeBuse, C., & Leynes, P. A. (2007). Conflict and criterion setting in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 2-17. doi:10.1037/0278-7393.33.1.2

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*(2), 452-478. doi:10.1037//0033-295X.108.2.452

Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal, 5,* 39-52.

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62*, 1-18. doi:10.1016/j.jml.2009.09.004

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 21-31. doi:10.1037/0278-7393.16.1.5

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8-20. doi:10.3758/bf03198438

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 5-16. doi:10.1037/0278-7393.16.1.5

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100*(3), 546-567. doi:10.1037/0033-295X.100.3.546

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*, 23-29. doi:10.1037/h0040561

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons, Inc

Groninger, L. D. (1976). Predicting recognition during storage: The capacity of the memory system to evaluate itself. *Bulletin of the Psychonomic Society, 7*(5), 425-428. doi:10.3758/bf03337236

Gronlund, S. D., & Ratcliff, R. (1989). Time Course of Item and Associative Information: Implications for Global Memory Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 846-858.

Heathcote, A. (2003). Item Recognition Memory and the Receiver Operating Characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1210–1230. doi: 10.1037/0278-7393.29.6.1210

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(6), 1306-1323. doi:10.1037/0278-7393.23.6.1306

Hoshino, Y. (1991). A bias in favor of the positive response to high-frequency words in recognition memory. *Memory & Cognition*, *19*(6), 607-616. doi:10.3758/bf03197156

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3), 306-340. doi:10.1037/0096-3445.110.3.306

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1534-1555. doi:10.1037//0278-7393.26.6.1534

Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. (1971). Recognition time for information stored in long-term memory. *Perception & Psychophysics, 10,* 8-14. doi: 10.3758/bf03205757

Link, S. W. (1975). The relative judgement theory of two-choice response time. *Journal Of Mathematical Psychology, 12*, 114-135. doi:10.1016/0022-2496(75)90053-X

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77-105. doi:10.1007/bf02291481

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition, 30*(4), 607-613. doi:10.3758/bf03194962

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252-271. doi:10.1037/0033-295X.87.3.252

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8*(3), 410-423. doi:10.1016/S0022-5371(69)80134-9

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*(4), 724-760. doi:10.1037/0033-295X.105.4.734-760

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, London: Chapman & Hall

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modelling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4*, 79-95. doi: 10.3758/BF03210778

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. *Journal Of Mathematical Psychology, 6*(2), 192-208. doi:10.1016/0022-2496(69)90002-9

Osth, A. F., Dennis, S., & Heathcote, A. (2017). Likelihood ratio sequential sampling models of recognition memory. *Cognitive Psychology*, *92*, 101-126. doi:10.1016/j.cogpsych.2016.11.007

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241-263. doi:10.1037/h0027272

Pinheiro, J. C., & Bates, D. M. (2000). *Statistics and Computing: Mixed-Effects Models in S and S-PLUS.* New York: Springer-Verlag

Rabbitt, P. (1969). Psychological refractory delay and response-stimulus interval duration in serial, choice-response tasks. *Acta Pscyhologica, 30,* 159-219. doi:10.1016/0001-6918(69)90051-1

Rabbitt, R., Cumming, G., & Vyas, S. (1978). Some errors of perceptual analysis in visual search can be detected and corrected. *Quarterly Journal of Experimental Psychology, 30*(2), 319-322. doi:10.1080/14640747808400679

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology*, *29*(4), 727-743. doi:10.1080/14640747708400645

Rabbitt, P., & Vyas, S. (1981). Processing a display even after you make a response to it. How perceptual errors can be corrected. *The Quarterly Journal of Experimental Psychology Section A, 33*(3), 223-239. doi:10.1080/14640748108400790

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59-108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356. doi: 10.1111/1467-9280.00067

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature, 461*(7261), 263-266. doi:10.1038/nature08275

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review, 12*(4), 573-604. doi:10.3758/bf03196750

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145-166. doi:10.3758/bf03209391

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal Of Mathematical Psychology, 32*(2), 135-168. doi:10.1016/0022-2496(88)90043-0

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50. doi:10.1037/0096-3445.117.1.34

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1379-1396. doi:10.1037/0278-7393.24.6.1379

Swensson, R. G. (1972). The elusive tradeoff: Speed vs accuracy in visual discrimination tasks. *Perception & Psychophysics, 12*, 16-32. doi:10.3758/bf03212837

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson (Eds.), *Organization Memory* (pp. 381-403). New York: Academic Press

Tulving, E. (1984). Précis of elements of episodic memory. *Behavioral and Brain Sciences, 7*(2), 223-238. doi: 10.1017/S0140525X0004440X

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*(4), 385-398. doi:10.1037/0003-066X.40.4.385

Tulving, E. (1995). Organization of memory: Quo vadis?. In M. S. Gazzaniga (Ed.), *The*

Cognitive Neurosciences (pp. 839-847). Cambridge: MIT Press

van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *Elife, 5,* e12192. doi:10.7554/eLife.12192

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 582-600. doi:10.1037//0278-7393.26.3.582

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1147-1166. doi:10.1037/0278-7393.30.6.1147

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13,* 37-58. doi:10.1080/00140137008931117

Vickers, D. (1979). *Decision Processes in Visual Perception*. London: Academic Press

Winograd, E., Cohen, C., & Barresi, J. (1976). Memory for concrete and abstract words in bilingual speakers. *Memory & Cognition, 4*(3), 323-329. doi:10.3758/bf03213184

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 681-690. doi:10.1037/0278-7393.18.4.681

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1341-1354. doi:10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*(6), 747-763. doi:10.3758/bf03211318

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and resposne speed. *Canadian Journal of Experimental Psychology*, *48*(4), 516-535. doi:10.1037/1196-1961.48.4.516

Appendix B

Participant Information Sheet

The Effects of Snap Decision Making on Episodic Recognition Memory

<u>Participant Information Sheet</u>

1. Invitation
You have been invited to participate in the research being conducted in partial fulfilment of a Bachelor of Psychological Science with Honours degree for Ellen-Jane Hickey under the supervision of Prof Andrew Heathcote.

2. What is the purpose of this study?
The aim of the study is to determine the effects of making a snap decision on recognition memory for words.

3. Why have I been invited to participate?
You have been invited to participate as you met the requirements for participation (i.e., you are 18 years or older). If you are a first year psychology undergraduate student you have replied to the SONA advertisement. If you are not a first year psychology student then you have seen an advertisement for the study and contacted Ellen-Jane with an expression of interest. The choice to participant is entirely voluntarily. There will be no consequences should you decide not to participate.

4. What will I be asked to do?
You will be asked to make a two-stage decision. First, you will be asked to study a list of words. You will then see another list of words, and for each word you will be asked to rapidly decide whether you have seen this word before (i.e., did you study it?) or not. Following this decision, you will then be asked to rate your confidence on your decision (high confidence, low confidence, or change of mind). This process will repeat for every word in each cycle of word lists you see. The entire experiment will be conducted here in the cognition lab and will take you approximately one hour to complete. You will be given these instructions in more detail along with practice trials prior to starting the experimental trials.

5. Are there any possible benefits from participation in this study?
There are no direct benefits to participants. The data collected in this research will provide further understanding on the effects of snap decision making on episodic recognition memory.

Directly following the completion of the experiment, first year psychology students will receive one hour of course credit via SONA for participating in the study, other participants will receive $15 from the experimenter as remuneration for your time.

6. Are there any possible risks from participation in this study?
There are no foreseeable risks involved with participating in this study. You may experience some fatigue. It is therefore recommended that you take a short break when prompted to.

7. What if I change my mind during or after the study?
If you wish to withdraw from the experiment you may do so without explanation up until you complete the experiment. It will not be possible to withdraw your data once you have completed the

experiment as all collected data is de-identified. This means that your name is not attached to any data that is collected, and thus cannot be identified. If you do not want your data to be collected, please withdraw before completing the experiment.

All de-identified data will also be made available to other researchers on an Open Science Framework (OSF) site. By signing the consent form you are consenting to your data being collected for the current study AND being made available on the OSF site. If you do not want your data made available to other researchers on the OSF site, please do not participate in the study, or if you change your mind once you have started, please withdraw before completing the experiment.

If you decide to withdraw part way through the experiment your incomplete data file will be deleted immediately.

8. What will happen to the information when this study is over?

Your consent form will be stored in a locked filing cabinet in the Cognition Lab. Electronic data files will be stored on a password protected server in the Cognition Lab. Hard copy and electronic files in the Cognition Lab will be kept for a minimum of five year before being destroyed. However, de-identified electronic data will be made available indefinitely to other researchers on an OSF site.

9. How will the results of the study be published?

The results of this study will be published as part of an honours thesis. Once published, the thesis will be accessible through the University of Tasmania library catalogue search function and the results of the study will be viewable. Alternatively, you may request a summary of the results by contacting Ellen-Jane (hickeyej@utas.edu.au). Results will be available in November. Participants will not be identifiable in the published results.

10. What if I have questions about this study?

If you have any questions or concerns about the study please contact either Ellen-Jane (hickeyej@utas.edu.au) or Andrew (Andrew.heathcote@utas.edu.au).

This study has been approved by the Tasmanian Social Sciences Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study, please contact the Executive Officer of the HREC (Tasmania) Network on +61 3 6226 6254 or email human.ethics@utas.edu.au. The Executive Officer is the person nominated to receive complaints from research participants. Please quote ethics reference number **H0016517**.

This information sheet is for you to keep. You will be given a separate consent form to read and sign should you choose to participate in the study.

Appendix C

Participant Consent Form

The Effects of Snap Decision Making on Episodic Recognition Memory

Participant Consent Form

1. I agree to take part in the research study named above.
2. I have read and understood the Information Sheet for this study.
3. The nature and possible effects of the study have been explained to me.
4. I understand that the study involves making a two-stage decision in a recognition memory task for words. I have read and understood the experiment instructions provided on the information sheet.
5. I understand that participation involves no foreseeable risks. I may experience some fatigue. I understand that it is recommended that I take short breaks when prompted to.
6. I understand that all research data will be securely stored on the University of Tasmania premises for a minimum of five years from the publication of the study results, and will then be destroyed.
7. I understand that my de-identified electronic data will be made available to other researchers on an Open Science Framework site indefinitely.
8. Any questions that I have asked have been answered to my satisfaction.
9. I understand that the researcher(s) will maintain confidentiality and that any information I supply to the researcher(s) will be used only for the purposes of the research.
10. I understand that the results of the study will be published so that I cannot be identified as a participant.
11. I understand that my participation is voluntary and that I may withdraw without explanation up until I complete the experiment.

    I understand that I will not be able to withdraw my data after completing the experiment as it will be non-identifiable.

Participant’s name: ____

Participant’s signature: ____

Date: ________________________

Statement by Investigator

☐ I have explained the project and the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation.

If the Investigator has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

☐ The participant has received the Information Sheet where my details have been provided so participants have had the opportunity to contact me prior to consenting to participate in this project.

Investigator's name: ____

Investigator's signature: ____

Date: ________________________

Appendix D

Data Analysis Output

D.1: First response generalised binomial probit linear mixed models output

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
O1.s	9	31901	31976	-15942	31883			
O1.ws	10	31893	31976	-15937	31873	10.055	1	0.0015

s= random subject effects; w=random item effects

ANOVA (O1.ws):

	χ^2	df	Pr(>Chisq)
NO	141.2177	1	<2.2e-16
WF	1.5833	1	0.20828
CC	6.2908	1	0.01214
NO:WF	30.2428	1	3.812e-08
NO:CC	4.0876	1	0.04320
WF:CC	0.3839	1	0.53551
NO:WF:CC	0.0807	1	0.77640

NO= new/old response; WF= word frequency; CC= word concreteness

Hit Rate ANOVA (O1.ws):

	χ^2	df	Pr(>Chisq)
WF	21.8747	1	2.91e-06
CC	10.4754	1	0.00121
WF:CC	0.3671	1	0.54459

NO= new/old response; WF= word frequency; CC= word concreteness

False Alarm Rate ANOVA (O1.ws):

	χ^2	df	Pr(>Chisq)
WF	8.5458	1	0.003463
CC	0.1517	1	0.696914
WF:CC	0.0621	1	0.803200

NO= new/old response; WF= word frequency; CC= word concreteness

D.1.1: Linear mixed effects model output for first response reaction times

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
RT1.s	10	5268.2	5351.1	-2624.1	5248.2			
RT1.ws	11	5269.8	5361.0	-2623.9	5247.8	0.3888	1	0.5329

s= random subject effects; w=random item effects

ANOVA (RT1.ws):

	χ^2	df	Pr(>Chisq)
NO	12.8983	1	0.0003289
WF	0.0369	1	0.8476350
CC	2.3669	1	0.1239293
NO:WF	1.4061	1	0.2357009
NO:CC	0.4197	1	0.5170855
WF:CC	0.6841	1	0.4081824
NO:WF:CC	0.0001	1	0.9919196

NO= new/old response; WF= word frequency; CC= word concreteness

D.1.2: Linear mixed effects model output for first response correctness reaction times

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
RT1.s	18	5111.9	5261.1	-2537.9	5075.9			
RT1.ws	19	5113.7	5271.1	-2537.8	5075.7	0.2304	1	0.6312

s= random subject effects; w=random item effects

ANOVA (RT1.ws):

	χ^2	df	Pr(>Chisq)
NO	12.2034	1	0.000477
WF	0.0313	1	0.859531
CC	1.8372	1	0.175284
C1	134.0645	1	< 2.2e-16
NO:WF	0.9843	1	0.321129
NO:CC	0.6965	1	0.403966
WF:CC	0.7785	1	0.377588
NO:C1	22.9170	1	1.692e-06
WF:C1	5.7314	1	0.016664
CC:C1	2.2222	1	0.136036
NO:WF:CC	0.0001	1	0.9919196
NO:WF:C1	6.9351	1	0.008452
NO:CC:C1	0.5896	1	0.442570
WF:CC:C1	0.0145	1	0.904085
NO:WF:CC:C1	0.1971	1	0.657043

NO= new/old response; WF= word frequency; CC= word concreteness, C1= first response correctness

D.2: Second response generalised binomial probit linear mixed models output

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
O2.s	9	34037	34111	-17010	34019			
O2.ws	10	34008	34091	-16994	33988	30.944	1	2.656e-08

s= random subject effects; w=random item effects

ANOVA (O2.ws):

	χ^2	df	Pr(>Chisq)
NO	4308.5023	1	<2.2e-16
WF	8.9125	1	0.0028183
CC	3.6317	1	0.0566886
NO:WF	84.8394	1	<2.2e-16
NO:CC	12.3611	1	0.0004384
WF:CC	0.0853	1	0.7701842
NO:WF:CC	0.0936	1	0.7596399

NO= new/old response; WF= word frequency; CC= word concreteness

Hit Rate ANOVA (O2.ws):

	χ^2	df	Pr(>Chisq)
WF	68.8930	1	<2.2e-16
CC	13.8572	1	0.0001972
WF:CC	0.1911	1	0.6620423

NO= new/old response; WF= word frequency; CC= word concreteness

False Alarm Rate ANOVA (O2.ws):

	χ^2	df	Pr(>Chisq)
WF	16.1944	1	5.716e-05
CC	0.9926	1	0.3191
WF:CC	0.0001	1	0.9938

NO= new/old response; WF= word frequency; CC= word concreteness

D.2.1: Linear mixed effects model output for second response reaction times

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
RT2.s	10	49215	49297	-24597	49195			
RT2.ws	11	49217	49308	-24597	49195	0	1	1

s= random subject effects; w=random item effects

ANOVA (RT2.ws):

	χ^2	df	Pr(>Chisq)
NO	53.3138	1	4.73e-13
WF	5.0059	1	0.025261
CC	0.2898	1	0.590367
NO:WF	2.5528	1	0.110102
NO:CC	9.8409	1	0.001707
WF:CC	1.1916	1	0.274999
NO:WF:CC	0.0523	1	0.819165

NO= new/old response; WF= word frequency; CC= word concreteness

D.2.2: Linear mixed effects model output for second response correctness reaction times

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
RT2.s	18	49140	49290	-24552	49104			
RT2.ws	19	49142	49300	-24552	49104	0	1	1

s= random subject effects; w=random item effects

ANOVA (RT2.ws):

	χ^2	df	Pr(>Chisq)
NO	53.0113	1	7.206e-14
WF	3.0921	1	0.078671
CC	0.1072	1	0.743319
C2	48.6399	1	3.076e-12
NO:WF	1.3896	1	0.236954
NO:CC	8.5947	1	0.003371
WF:CC	1.2020	1	0.272922
NO:C2	32.7423	1	1.052e-08
WF:C2	0.5211	1	0.470394
CC:C2	0.2668	1	0.605476
NO:WF:CC	0.0954	1	0.757434
NO:WF:C2	4.0930	1	0.043061
NO:CC:C2	2.6505	1	0.103518
WF:CC:C2	0.5694	1	0.450515
NO:WF:CC:C2	0.7610	1	0.383013

NO= new/old response; WF= word frequency; CC= word concreteness, C2= second response correctness

D.3: Generalised binomial probit linear mixed models output for combined first and second response

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
O12.s	17	67482	67634	-33724	67448			
O12.ws	18	67119	67280	-33541	67083	365.14	1	<2.2e-16

s= random subject effects; w=random item effects

ANOVA (O12.ws):

	χ^2	df	Pr(>Chisq)
R12	7.6167	1	0.0057830
NO	5276.6677	1	<2.2e-16
WF	5.6228	1	0.0177287
CC	6.1707	1	0.0129882
R12:NO	563.2737	1	<2.2e-16
R12:WF	2.2249	1	0.1358035
NO:WF	105.9609	1	<2.2e-16
R12:CC	0.0865	1	0.7687260
NO:CC	0.0281	1	0.8669344
WF:CC	0.0281	1	0.8669344
R12:NO:WF	9.3421	1	0.0022394
R12:NO:CC	1.2918	1	0.2557175
R12:WF:CC	0.4801	1	0.4883791
NO:WF:CC	0.0102	1	0.9195010
R12:NO:WF:CC	0.2489	1	0.6178620

R12= combined response 1 and response 2 data; NO= new/old response; WF= word frequency; CC= word concreteness

Hit Rate ANOVA (O12.ws):

	χ^2	df	Pr(>Chisq)
R12	231.3853	1	<2.2e-16
WF	54.4863	1	1.565e-13
CC	14.0572	1	0.0001773
R12:WF	10.9142	1	0.0009543
R12:CC	0.3456	1	0.5566034
WF:CC	0.0001	1	0.9938831
R12:WF:CC	0.6947	1	0.4045748

R12= combined response 1 and response 2 data; NO= new/old response; WF= word frequency; CC= word concreteness

False Alarm Rate ANOVA (O12.ws):

	χ^2	df	Pr(>Chisq)
R12	389.8509	1	<2.2e-16
WF	14.6428	1	0.0001299
CC	0.1048	1	0.7461302
R12:WF	1.3272	1	0.2493024
R12:CC	1.1096	1	0.2921751
WF:CC	0.0171	1	0.8959944
R12:WF:CC	0.0297	1	0.8631760

R12= combined response 1 and response 2 data; NO= new/old response; WF= word frequency; CC= word concreteness

D.4: Response consistency reaction time output

Model fits:

	df	AIC	BIC	Loglik	deviance	χ^2	χ^2df	Pr(>Chisq)
RTCON.s	18	45411	45560	-22688	45370			
RTCON.ws	19	45413	45571	-22688	45375	0.0052	1	0.9423

s= random subject effects; w=random item effects

ANOVA (RTCON.ws):

	χ^2	df	Pr(>Chisq)
NO	66.2851	1	3.902e-16
WF	7.5161	1	0.006115
CC	2.2495	1	0.133654
CON	4040.9612	1	<2.2e-16
NO:WF	5.5218	1	0.018781
NO:CC	6.2154	1	0.012664
WF:CC	0.5936	1	0.441022
NO:CON	31.9008	1	1.622e-08
WF:CON	0.0043	1	0.947671
CC:CON	0.4060	1	0.524011
NO:WF:CC	0.0041	1	0.948893
NO:WF:CON	2.1311	1	0.144336
NO:CC:CON	0.4443	1	0.505069
WF:CC:CON	1.0044	1	0.316238
NO:WF:CC:CON	0.2746	1	0.600265

NO= new/old response; WF= word frequency; CC= word concreteness, CON=second response consistency